

Anotación semántica de *Web Feature Services*

Víctor Saquicela, Luis M. Vilches-Blázquez, Oscar Corcho

Ontology Engineering Group, Dpto. Inteligencia Artificial,
Fac. Informática, Univ. Politécnica de Madrid,
Av. Montepríncipe s/n, 29660 Boadilla del Monte, Madrid, España
{vsaquicela, lmvilches, ocorcho}@fi.upm.es

Resumen

Los servicios WFS permiten recuperar información geográfica de *features*. La anotación semántica de estos servicios proporciona beneficios en tareas relacionadas con el descubrimiento, composición y mediación de servicios. La principal contribución de este trabajo consiste en proponer un proceso para la automatización de la anotación semántica de servicios WFS. Para estas anotaciones se utilizan diversos recursos, tales como: ontologías y conjuntos de datos presentes en la nube de *Linked Data*.

Palabras clave: *Web Feature Services*, anotación semántica, ontologías, *Linked Data*.

1 Introducción

Actualmente, muchas organizaciones, tanto públicas como privadas, están generando servicios Web conforme a las especificaciones del *Open Geospatial Consortium* (OGC). Estos servicios, normalmente, son recopilados en sitios Web o catálogos donde son descritos mediante una URL y una serie de elementos de metadatos descriptivos. De esta manera, las descripciones de estos servicios son únicamente sintácticas, ya que su descripción se recoge generalmente en ficheros de tipo HTML, texto y/o XML. Esta situación permite pensar en actividades relacionadas con la descripción (anotación) semántica de estos servicios. Esto permitiría el enlazado de descripciones significativas a los servicios y datos geoespaciales.

Tradicionalmente, los enfoques propuestos para abordar el problema de la anotación semántica de servicios Web se han centrado en definir formalmente las

descripciones de los servicios. Recientemente, estos enfoques se están utilizando para describir semánticamente servicios OGC, ya que estas descripciones proporcionan beneficios en tareas relacionadas con el descubrimiento, composición y mediación de servicios. Sin embargo, las tareas relacionadas con la anotación todavía requieren una significativa intervención humana, debido a que el proceso es completamente manual.

La principal contribución de este trabajo consiste en proponer un proceso automático de anotación semántica para servicios *Web Feature Services* (WFS). Para ello, se tratan, principalmente, 2 aspectos: 1) se proporciona una descripción sintáctica de servicios WFS para realizar invocaciones automáticas y 2) se interpretan y enriquecen las descripciones de estos servicios (WFS) mediante anotaciones semánticas. Para estas anotaciones se utilizan diversos recursos, tales como: ontologías y conjuntos de datos presentes en la nube de *Linked Data*.

En este artículo se presenta la descripción del proceso de anotación semántica. Primero se hace una breve descripción de los servicios WFS y los trabajos relacionados, a continuación se describe el enfoque utilizado en el proceso de anotación. Por último se presenta una serie de conclusiones y trabajos futuros.

2 Servicios WFS y trabajos relacionados

Esta sección recoge una descripción de las principales características de los servicios WFS, así como diversas propuestas relacionadas con la anotación semántica en el contexto geoespacial.

2.1 *Web Feature Services*

WFS es un servicio Web que permite consultar y recuperar datos vectoriales y la información alfanumérica ligada a los mismos. Además, este servicio permite realizar consultas tanto espaciales como alfanuméricas.

Los servicios WFS soportan tres operaciones: *GetCapabilities*, *DescribeFeatureType* y *GetFeature*. A continuación se recoge una breve descripción de cada uno.

GetCapabilities. Esta operación describe las funcionalidades y los datos ofrecidos por un servicio determinado. Esto incluye un listado de *features* geográficos. A través del *GetCapabilities* se pueden obtener los metadatos del servicio e

identificar el tipo de información que posee. Esta operación proporciona un documento (XML) que describe las características del servicio. Este documento está compuesto de *Service*, *Capabilities* y *FeatureTypeList*.

DescribeFeatureType. El propósito de esta operación es recuperar información de la estructura de un *feature*. Específicamente, la operación *DescribeFeatureType* recupera un listado de propiedades y atributos para un *feature* determinado o un listado de todos los tipos de *feature* disponibles.

GetFeature. La operación *GetFeature* recupera los datos existentes (instancias), incluyendo la geometría y los valores de los atributos.

Para obtener detalles más específicos de estas operaciones y sus características se recomienda consultar¹.

2.2 Trabajo relacionado

Con respecto a los enfoques de anotación existentes, en [1] los autores definen métodos semánticos para construir servicios Web geoespaciales utilizando el enfoque REST² (*Representational State Transfer*). En este trabajo se presentan dos métodos para realizar correspondencias entre recursos geoespaciales y el mundo semántico de acuerdo a los principios REST. En [2] se propone un enfoque de anotación semántica utilizando metadatos geográficos y ontologías para describir datos geoespaciales heterogéneos. En este trabajo se utilizan ficheros RDF/XML y el esquema de metadatos *Federal Geographic Data Community* (FGDC) para enlazar términos ontológicos y los metadatos.

Asimismo, ESRI³ propone una especificación de servicios geoespaciales basada en los principios REST. Esta especificación provee una forma estándar de comunicación para clientes Web con los Sistema de Información Geográfica (SIG) a través de la tecnología REST. Los clientes envían peticiones al servidor utilizando URL estructuradas y el servidor responde con imágenes, textos u otros recursos que satisfacen la petición.

En [3] se presenta un enfoque para realizar automáticamente anotaciones semánticas de datos geoespaciales. En este trabajo destaca el rol de las relaciones

¹ <http://www.opengeospatial.org/standards/wfs>

² <http://searchsoa.techtarget.com/definition/REST>

³ <http://www.esri.com/library/whitepapers/pdfs/geoservices-rest-spec.pdf>

espaciales para extraer información implícita a nivel de instancia. Además sugiere la aplicación de métodos de análisis espacial para extraer información de relaciones espaciales útiles para las anotaciones.

Finalmente, en [4] se define un conjunto de recomendaciones para añadir semántica a diversos estándares y/o especificaciones geoespaciales existentes, tales como: GML, KML, esquema GML, OWS, etc. En este trabajo se sugieren tres niveles de anotación: 1) Anotación de los metadatos de los recursos (OWS *capabilities*), 2) anotación del modelo de datos (esquema GML) y 3) anotación de los datos de las entidades (archivos GML).

3 Un enfoque automático para la anotación semántica de servicios WFS

En esta sección se presenta el enfoque propuesto para automatizar el proceso de anotación semántica de servicios WFS. Una visión general de la arquitectura propuesta se muestra en la figura 1. El sistema desarrollado está compuesto de 3 componentes: 1) componente de invocación y registro, 2) repositorio, donde se almacenan las descripciones de los servicios, y 3) componente de anotación semántica.

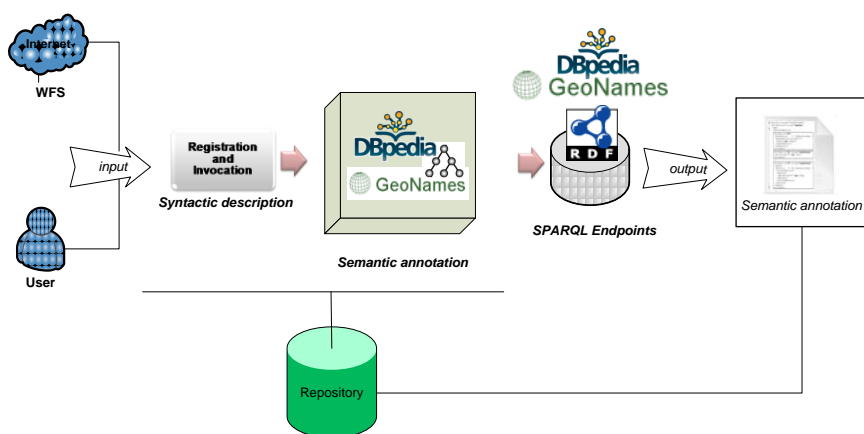


Figura 1. Arquitectura para anotación semántica de servicios WFS

Para demostrar la validez del sistema desarrollado se ha seleccionado un servicio WFS que será utilizado para aportar ejemplos en las diferentes secciones.

Servicio 1.

http://apps.who.int/tools/geoserver/wfs?REQUEST=GetCapabilities&SERVICE=WFS

Este servicio es desarrollado por *World Health Organization* y contiene información sanitaria controlada por el sistema de las Naciones Unidas.

A continuación, se describen brevemente los componentes del sistema ilustrando esta descripción con ejemplos.

3.1 Descripción sintáctica de servicios WFS: Registro de la invocación en el repositorio

Como se mencionó anteriormente, los servicios WFS son normalmente registrados en sitios Web o catálogos utilizando URL y una serie de metadatos. Considerando este hecho, el sistema propuesto en el contexto de este trabajo toma como entrada la URL de un servicio WFS que esté disponible. Una vez que la URL se agrega, el sistema invoca al servicio WFS y analiza la respuesta para obtener información de las principales características del servicio. Como respuesta a esta petición (*GetCapabilities*), el sistema obtiene un fichero XML que describe las *capacidades* del servicio.

En el tabla 1 se visualiza una sección de la respuesta (XML) que genera la operación *getCapabilities* sobre el WFS considerado (Servicio 1). Esta respuesta se utiliza para obtener información de las características que describen al referido servicio.

Tabla 1: Resultado de la operación GetCapabilities (Servicio 1)

```
<wfs:WFS_Capabilities
.....
xmlns:ows="http://www.opengis.net/ows"
.....
<FeatureType xmlns:WHO="http://www.who.int/">
  <Name>WHO:efs2008_countrydata</Name>
  <Title>EpiFact Sheets 2008 Data</Title>
  <Abstract>Generated from WHOGeodata</Abstract>
  <ows:Keywords>
    <ows:Keyword>efs_data_for_webservice</ows:Keyword>
    <ows:Keyword>WHOGeodata</ows:Keyword>
  </ows:Keywords>
  <DefaultSRS>urn:x-ogc:def:crs:EPSG:4326</DefaultSRS>
  <ows:WGS84BoundingBox>
    <ows:LowerCorner>
      -527.400749390449 -545.3687536171334
```

```

        </ows:LowerCorner>
        <ows:UpperCorner>
            586.4983522346869 568.5303480080025
        </ows:UpperCorner>
    </ows:WGS84BoundingBox>
</FeatureType>
    .....
</wfs:WFS Capabilities>

```

Cuando se necesita el listado de todos los tipos de *feature*, la operación utilizada es la siguiente (ver tabla 2):

Tabla 2: URL para recuperar el listado de feature

```

http://apps.who.int/tools/geoserver/wfs?REQUEST=DescribeFeatureType&SERVICE=WFS

```

Con esta operación se obtiene un listado de 44 diferentes tipos de *features* (por ejemplo, *WHO:Admin_world*, *WHO:world_countries*, *WHO:africa_airport*, etc.).

Cuando se requiere un tipo específico de *feature*, se utiliza la operación *DescribeFeatureType*. En este caso la operación GET aplicada es (ver tabla 3):

Tabla 3: URL para recuperar el listado de parámetros de un feature

```

http://apps.who.int/tools/geoserver/wfs?REQUEST=DescribeFeatureType&TYPENAME=WHO:FLUID_DEV_GIM3_MONITORING_GEOMETRY_MEASURES

```

En la respuesta XML recogida en el *tabla 4*, los parámetros presentes no se describen utilizando palabras que puedan ser entendidas por usuarios y por maquinas. Por ejemplo, el parámetro COUNTRY_ID está compuesto de dos palabras, COUNTRY e ID. Este tipo de parámetros se procesan utilizando métricas de comparación como se describe en la siguiente sección.

Tabla 4. Operación DescribeFeatureType. Listado de parámetros

```

<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  <xsd:complexType
name="FLUID_DEV_GIM_MONITORING_GEOMETRY_MEASURESType">
  <xsd:complexContent>
    <xsd:extension base="gml:AbstractFeatureType">
      <xsd:sequence>
        <xsd:element
m\label{subsub:DescribeFeatureType}axOccurs="1"

```

```

        minOccurs="1" name="COUNTRY_ID" nillable="true"
type="xsd:int"/>
        <xsd:element maxOccurs="1" minOccurs="1"
        name="REGION_ID" nillable="true"
type="xsd:int"/>
        <xsd:element maxOccurs="1" minOccurs="1"
        name="COUNTRY_NAME" nillable="true"
type="xsd:string"/>
        <xsd:element maxOccurs="1" minOccurs="0"
        name="ISO_CODE" nillable="true"
type="xsd:string"/>
        <xsd:element maxOccurs="1" minOccurs="0"
        name="REGION_CODE" nillable="true"
type="xsd:string"/>
        .....
    </xsd:sequence>
</xsd:extension>
</xsd:complexContent>
</xsd:complexType>
.....
</xsd:schema>

```

Tras la obtención de las características del *feature* WHO:FLUID_DEV_GIM_MONITORING_GEOMETRY_MEASURES (*feature* que pertenece al Servicio 1), utilizando la operación *DescribeFeatureType* se recuperan todas las propiedades definidas para las instancias del mencionado *feature*. Entre esta información aparece la geometría y los valores de los atributos asociados.

Una vez conocida la estructura del *feature*, se procede a invocar a la operación *GetFeature*, que retorna un fichero GML de un *feature* geográfico asociado a geometrías y atributos (ver tabla 5).

Tabla 5: Operación GetFeature

```

http://apps.who.int/tools/geoserver/wfs?REQUEST=GetFeature&
&TYPENAME=WHO:FLUID_DEV_GIM_MONITORING_GEOMETRY_MEASURES
&FILTER=<Filter>
<PropertyIsEqualTo><PropertyName>COUNTRY_ID</PropertyName>
<Literal>9</Literal></PropertyIsEqualTo></Filter>

```

Por ejemplo, para recuperar las características del *feature* WHO:FLUID_DEV_GIM_MONITORING_GEOMETRY_MEASURES, el sistema utiliza la operación *GetFeature* con el parámetro COUNTRY_ID y valor 9. Esto permite

utilizar filtros para la recuperación de la información. La respuesta generada se visualiza en el tabla 6.

Tabla 6: Operación GetFeature para recuperar un listado de instancias

```
<wfs:FeatureCollection "
.....
  <gml:featureMembers>
    <WHO:FLUID_DEV_GIM_MONITORING_GEOMETRY_MEASURES

gml:id="FLUID_DEV_GIM_MONITORING_GEOMETRY_MEASURES.204">
  <WHO:COUNTRY_ID>9</WHO:COUNTRY_ID>
  <WHO:REGION_ID>6</WHO:REGION_ID>
  <WHO:COUNTRY_NAME>Australia</WHO:COUNTRY_NAME>
  <WHO:ISO_CODE>AUS</WHO:ISO_CODE>
  <WHO:LOCATION_ID>46</WHO:LOCATION_ID>
  <WHO:COUNTRY_GEOMETRY>
    <gml:MultiSurface                                srsName="urn:x-
ogc:def:crs:EPSG:4326">
      <gml:surfaceMember>
        <gml:Polygon>
          <gml:exterior>
            <gml:LinearRing>
              <gml:posList>
                -12.008417859307 143.11931505333
                -14.253175795324 144.516width05332463
                -19.017050970651 146.36174318313
                .....
              </gml:posList>
            </gml:LinearRing>
          </gml:exterior>
        </gml:Polygon>
      </gml:surfaceMember>
      .....
    </gml:MultiSurface>
  </WHO:COUNTRY_GEOMETRY>
  .....
</gml:featureMembers>
</wfs:FeatureCollection>
```

El siguiente paso en el sistema desarrollado conlleva la manipulación automática de los ficheros XML asociados a cada servicio WFS. Para esto, el sistema utiliza la

API Geotools⁴ para ejecutar la invocación del servicio WFS y determinar si está disponible o no. El proceso de ejecución se realiza de la siguiente manera: primero, el sistema toma como entrada la URL del servicio WFS. Seguidamente, el sistema invoca al servicio que traduce la invocación WFS en consultas específicas para el servicio y finalmente la respuesta se procesa utilizando GeoTools. Esta herramienta permite navegar a través de las características del servicio WFS.

La invocación de estos servicios utilizando GeoTools permite recuperar la respuesta del servicio en formato GML. Esta respuesta describe sintácticamente el servicio, cuya descripción se realiza conforme a un modelo diseñado para este sistema (ver figura 2). El modelo se basa en las ontologías WSMO-lite y POSM. Ontologías que permiten describir la estructura de un servicio web. Asimismo la información obtenida se almacena en una base de datos utilizando un modelo relacional. El modelo relacional está implementado como una base de datos, diseñada específicamente para almacenar descripciones sintácticas de servicios WFS. Se ha optado por esta forma de almacenamiento de la información para incrementar la eficiencia en la recuperación de los servicios WFS.

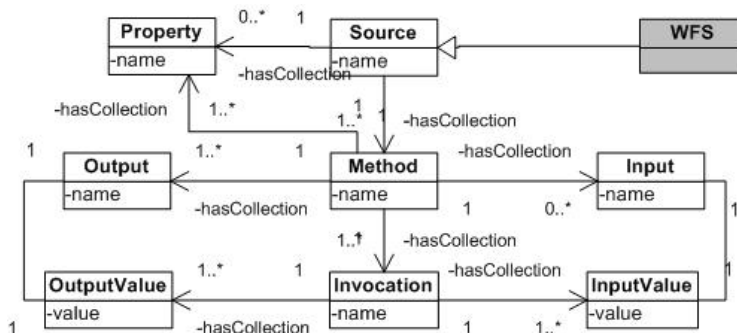


Figura 2. WFS. Un modelo para la descripción de servicios WFS

Un listado de todas las descripciones sintácticas obtenidas se recogen en la siguiente URL⁵.

Finalmente, el sistema manipula estos parámetros para anotar semánticamente el contenido del servicio WFS.

⁴ Geotools es una librería escrita en código Java que provee métodos estándares para la manipulación de datos geoespaciales y está basada en las especificaciones *Open Geospatial Consortium* (OGC). <http://geotools.org/>

⁵ <http://castor.dia.fi.upm.es/WFSannotation>

3.2 Anotación semántica de servicios WFS

Tras la descripción sintáctica, descrita en la sección anterior, se procede a la asignación de semántica a los servicios WFS. El punto de partida del proceso de anotación semántica es el listado de parámetros obtenidos previamente en la descripción sintáctica. Los servicios WFS están sintácticamente descritos utilizando los parámetros de entrada/salida y las propiedades de cada una de las operaciones (*GetCapabilities*, *DescribeFeatureType* y *GetFeature*).

En el proceso de anotación semántica de estos parámetros, el sistema propuesto sigue un enfoque heurístico que combina servicios externos y recursos semánticos (ver figura 3). A continuación, se describen los principales componentes del proceso de anotación semántica.

Utilización de recursos semánticos en el proceso de anotación

En este punto, el listado de parámetros del servicio WFS, obtenidos previamente, se utiliza para consultar recursos semánticos y recuperar resultados asociados a cada parámetro. El proceso seguido y los elementos utilizados se describen a continuación:

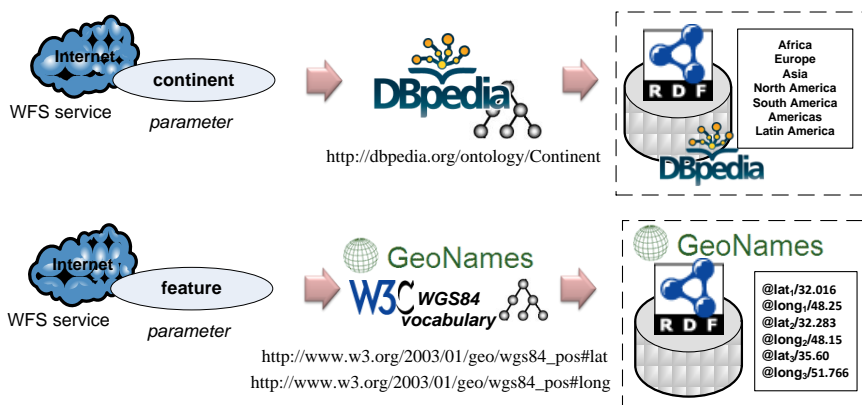


Figura 3. WFS. Proceso de anotación semántica

- Las ontologías utilizadas en este trabajo son:

- DBpedia⁶. Una ontología de dominio que ha sido manualmente creada basada en Wikipedia.
 - GeoNames⁷. Una ontología que añade información semántica geoespacial para la Web.
- El sistema recupera de estas dos ontologías todas las clases cuyos nombres coinciden con cada parámetro del servicio WFS. En este proceso de emparejamiento se utiliza una combinación de métricas de comparación de cadenas (*Jaro*, *JaroWinkler*, *Levenshtein*) para comparar los parámetros del servicio WFS con los conceptos de las ontologías (clases y propiedades). Esta propuesta permite, por ejemplo, el emparejamiento de cadenas, tales como *countryName*, *country_name* o *country*.
- Si el sistema obtiene correspondencias en el proceso de emparejamiento, este usa los conceptos (clases) de las dos ontologías mencionadas para recuperar ejemplos (instancias de conceptos) desde los *SPARQL Endpoint* relacionados con DBpedia y GeoNames. La información resultante se sugiere automáticamente al sistema y queda registrada como posibles valores para el parámetro correspondiente. Cuando el parámetro se empareja más de una vez con los elementos (clases) de la ontología, el sistema solo considera los conceptos que tienen información (instancias) y las que no tienen información son descartadas automáticamente.
- A continuación, el sistema busca correspondencias entre los parámetros del servicio WFS y las propiedades de las ontologías. Si el sistema obtiene algunas correspondencias, este utiliza las propiedades de las ontologías individualmente para recuperar información de los *SPARQL Endpoint* de DBpedia o GeoNames.
- Finalmente, con las clases y propiedades de las ontologías utilizadas el sistema llama a los *SPARQL Endpoint* de DBpedia y GeoNames para recuperar valores (instancias) para estos elementos. Con esto se obtiene un listado de posibles valores para cada parámetro del servicio.

4 Conclusiones y trabajos futuros

En este trabajo se propone un enfoque para realizar anotaciones semánticas de servicios WFS. Este proceso se implementa en un sistema que toma en cuenta las ontologías de DBpedia, GeoNames y sus respectivos *SPARQL Endpoint*. El

⁶ <http://dbpedia.org/>

⁷ <http://geonames.org>

sistema utiliza una combinación de estos recursos para descubrir el significado de cada uno de los parámetros recogidos en los servicios WFS. Además, para demostrar los resultados ofrecidos por este trabajo y guiar la explicación del proceso propuesto se ha utilizado un servicio WFS como ejemplo.

El trabajo futuro será orientará hacia la utilización de recursos externos como, por ejemplo, servicios de sugerencias y sinónimos. La utilización de estos servicios incrementará las posibilidades de anotación semántica de los servicios WFS. Además, para demostrar la viabilidad del sistema desarrollado se planea realizar una experimentación exhaustiva con una muestra representativa de servicios WFS.

Agradecimientos. Este trabajo ha sido financiado por el proyecto España Virtual (CENIT-2008-1030), auspiciado por el Centro Nacional de Información Geográfica y CDTI, enmarcado en el contexto del programa Ingenio 2010.

Referencias

- [1] Yao Gang. A research on semantic geospatial web service based rest. In Computer Science-Technology and Applications. IFCSTA '09. International Forum on, volume 1, pages 208–210. (2009)
- [2] Sidney Roberto Sousa. A semantic approach to describe geospatial resources. In Proceedings of the ER 2009 Workshops (CoMoL, ETheCoM, FP-UML, MOST-ONISW, QoIS, RIGiM, SeCoGIS) on Advances in Conceptual Modeling - Challenging Perspectives, ER '09, pages 327–336, Springer-Verlag. (2009).
- [3] Eva Klien y Michael Lutz. The role of spatial relations in automating the semantic annotation of geodata. In In COSIT, pages 133–148 (2005).
- [4] Patrick Maue, Sven Schade, y Philippe Duchesne. Semantic Annotations in OGC Standards. Open Geospatial Consortium (OGC), July 2009. Disponible en: http://portal.opengeospatial.org/files/?artifact_id=34916.
- [5] Kahan, J. Marja-Riitta Koivunen, Eric Prud'Hommeaux, y Ralph R. Swick. Annotea: An open rdf infrastructure for shared web annotations. Proceeding WWW '01 Proceedings of the 10th international conference on World Wide Web (2001).
- [6] Giorgos Stamou, Jacco Van, Jeff Z. Pan, Guus Schreiber, y John R. Smith. Multimedia annotations on the semantic web. IEEE MultiMedia, 13:86–90, (2006).